# A Portable Modular System for Automatic Acquisition of 3D Objects

Stefano Ferrari[1] and Nunzio Alberto Borghese[2, §]

[1] Centro di Bioingegneria, Fondazione ProJuventute, and
Dipartimento di Ingegneria Elettronica e Informatica, Politecnico di Milano, Italy.

[2] Laboratory of Human Motion Study and Virtual Reality,
Istituto Neuroscienze e Bioimmagini – CNR. Milano, Italy.

## Abstract

*A modular system which is able to reconstruct the 3D surface of an object is presented here. It has a three level architecture. The first level is devoted to the acquisition of a set of 3D points over the surface (digitisation), the second level constructs the 3D surface in the form of a mesh, filtering the measurement noise. In the third level a bitmap of the surface, obtained from a snapshot, is projected over the 3D mesh to obtain a highly realistic 3D model. This instrument improves the commercial available scanners in two main aspects. The digitiser proves highly flexible and it can easily accommodate objects of different dimension. The construction of the mesh and the filtering of the digitisation noise is carried in a single step through an algorithm which can be parallelised to work in real time. When the spot detection will be transferred to a standard graphic board and the mesh construction over a dedicated (FPGA) board, this instrument shall be seen as a standard input device of next generation graphical workstations.*

## 1 . Introduction

Virtual 3D models are required by an increasing number of applications ranging from basic image processing to video conferencing, constructive and plastic surgery, 3D fax, reversed engineering and 3D CAD. A host of devices (3D scanners), which provide these 3D models, have come to the market in the last few years [8].

Although ultrasound [11] or mechanical (e.g. Microscribe$^{TM}$) devices are available, optical technology is preferred because it allows a high resolution and it does not require any contact with the surface. The gold-standard is represented by the Cyberware$^{TM}$ scanners which are suitable to most applications. However, apart their very high cost, they have two drawbacks: a complex structure, inside which the object is placed, has to be set up and only objects within limited sizes range can be digitised.

In this paper, a 3D scanner, which combines flexibility and accuracy and it is easily portable is described. It is subdivided into three main modules. The first module is devoted to sampling a sufficient number of 3D points over the surface (digitisation); the second module transforms the set of points (range data) into a 3D surface, cleaning the noise introduced by the digitisation process; and the third module applies to the 3D model the texture, in the form of bitmap taken from a snapshot of the surface. The system has been widely tested in the 3D reconstruction of Human faces. This is a particularly difficult task because face's spatial frequency content is highly variable; moreover, small head motion during the digitisation process produces a measurement error which adds to measurement noise.

## 2 . Data acquisition

3D digitisation is carried out by the Autoscan system, described in [1]. This is constituted of a commercial laser pointer of 5mW, a pair of CCD video cameras, which provide an image of 256×256 pixels with a frame rate of 100 Hz, a real-time image processor and a host computer. In this version of the system, the Elite system image processor [6] has been adopted. This is widely used in automatic human motion analysis and it is able to recognise in real-time, spherical markers attached to repere points on the moving subject. Here, it is used to detect, on the subject, the projected laser spot, which constitutes a "virtual" marker. It can work in real-time thanks to the Elite cross-correlation between a mask, template of the spot, and the image carried out in real-time through a custom board. Moreover, it achieves a high

§ Corresponding author: Laboratory of Human Motion Study and Virtual Reality, INB - CNR via f.lli Cervi 93 - 20090 Segrate (MI), Italy, email borghese@inb.mi.cnr.it

0-7803-5276-9/99/$10.00 © 1999 IEEE

SNR, which allows detecting the laser spot also in outdoor conditions, with an experimental accuracy above 0.1 pixels. In the next version, the spot detection will be transferred from the Elite to a board (e.g. FPGA) in the host computer.

Surface digitisation is carried out moving the laser pointer manually. To help directing the laser beam, a real-time feedback is provided on the host PC monitor. This scanning procedure offers the great advantage to increase the number of sampled points in those regions where the surface is more variable. A typical ensemble of digitised points is reported in Figure 1a.

At the same time, through a standard CCD camera a snapshot of the face is acquired (Fig. 1c).

data points collected as the set $\{(P_j, z(P_j)) \mid P_j = (x_j, y_j) \in R^2, z: R^2 \rightarrow R\}$.

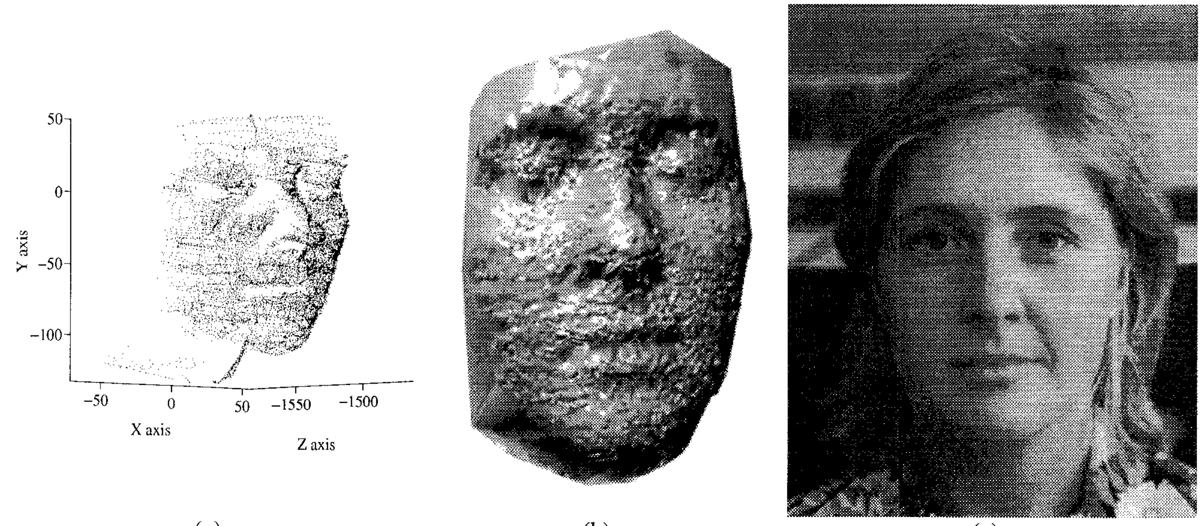
The network combines the output of many simple units to achieve the reconstruction of a complex surface. In particular, a HRBF Network is composed of radially symmetric Gaussian units:

$$G(P;\mu \mid \sigma) = \frac{1}{\sqrt{\pi}\sigma} e^{-\frac{(P-\mu)^2}{\sigma^2}} \qquad (1)$$

where $P, \mu \in R^2$ and $\sigma \in R$. In the HRBF model, the units are organised in layers, where each layer is composed of equally spaced Gaussians, which have the same standard deviation, $\sigma$. Therefore the surface, $z(P)$, is constructed by adding the contribution of few grids of Gaussian functions, where each grid operates at a certain scale (or cut-off frequency). Given a set, $\alpha$, of parameters which characterise the HRBF network, the actual shape of the reconstructed surface (i.e. the output of the net), $S(P|\alpha)$, is:

$$S(P \mid \alpha) = \sum_{l=1}^{L} \sum_{k=1}^{M_l} w_{kl} G(P; P_{kl} \mid \sigma_l) \qquad (2)$$

where $L$ is the number of grids and $M_l$ is the number of Gaussian units in the $l^{th}$ grid. $w_{kl}$ is the weight associated to the $k^{th}$ Gaussian in the $l^{th}$ grid and $P_{kl}$ is its position. $\sigma_l$ is the standard deviation of all the Gaussians in the $l^{th}$ grid which determines the filtering scale of that grid. The parameters $\alpha = \{L, M, \{w_{kl}\}, \{P_{kl}\}, \{\sigma_l\}\}$ determine the



(a) (b) (c)

***Figure 1. The ensemble of 12,641 digitised points is reported in (a). The construction of a mesh by simply connecting these points produces a noisy reconstruction (b). In (c) the face snapshot used for texture mapping is reported.***

## 3 . Surface reconstruction

Due to measurement noise, a direct tessellation of the data points, $\{P_j = (x_j, y_j, z_j)\}$, obtained by simply connecting them, would produce an undesirable wobbling surface (Fig. 1b), and the need of some sort of filtering in the reconstruction, is evident. This is achieved here through a particular model, the Hierarchical Radial Basis Function Network (HRBF), proposed originally in the connectionist domain [2]. We assume here that the surface could be represented as a function $S$: $R^2 \rightarrow R$. This assumption is motivated by the acquisition set-up used. Under this hypothesis, it is more convenient to reframe the actual shape of $S(P|\alpha)$. Although these could be determined with gradient descent algorithms on a regularising cost function [10], the offered solution is often very poor in this case and different solution schemas have been explored [3, 7, 9]. HRBF, in particular, offers a very fast solution as the determination of the parameters is performed with local operations carried out on the data points and it is particularly suitable to this application.

Each grid of the HRBF model realises a linear filter which is able to reconstruct the surface up to a certain scale, determined by $\sigma_l$. It can be shown [2] that $\sigma_l$ and the spacing between two consecutive Gaussians on the same grid, $\Delta P_l$, are related with:

$$\sigma_l = 1.465\, \Delta P_l \qquad (3)$$

This relationship is obtained accepting a maximum attenuation in the Pass Band of -3dB and a minimum attenuation in the Stop Band of -40dB. Different attenuation values lead to a different proportionality constant between $\Delta P_l$ and $\sigma_l$. To apply the Gaussian filter, to the sampled data, these should be equally sampled in correspondence of the grid crossings. That is the set $\{z_{kl}\} = \{z(P_{kl})\}$ should be available. Unfortunately this is not the case here; but, since many points are usually digitised (surface oversampling), a reliable estimate of $z(P_{kl})$ can be obtained through the following MAP estimator:

$$\tilde{z}(P_{kl}) = \frac{\sum\limits_{P_r \in A(P_{kl})} z(P_r)\, e^{-\frac{\|P_{kl}-P_r\|^2}{\sigma_w^{\,2}}}}{\sum\limits_{P_r \in A(P_{kl})} e^{-\frac{\|P_{kl}-P_r\|^2}{\sigma_w^{\,2}}}} \qquad (4)$$

where $A(P_{kl})$ is the Receptive field associated to the $k$-th Gaussian in the grid $l$. It is set, somehow arbitrary, as the circular region included in $P_k \pm \Delta P_l$. This estimation is carried out locally on the input space and it can be parallelised to achieve quasi real-time processing.

The grid filter can be written as:

$$S_r(P) = \sum_{k=1}^{M_l} \tilde{z}(P_{kl}) G(P; P_{kl} \mid \sigma_l) \Delta P_l^{\,2} \qquad (5)$$

Comparing (5) with (2) it can be demonstrated that the parameters $\{w_{kl}\}$ can be obtained simply as $\tilde{z}(P_{kl})\,\Delta P_l^2$, [2].

If only one grid are adopted, a large drawback is introduced: the Guassians scale should be small enough to resolve the finest details. This requires a very dense packing of the Gaussians also in those regions where the details can be resolved at a coarser scale, causing a waste of resources and overfitting in those space regions. A better solution would be to adaptively allocate the Gaussian units, with an adequate scale, in the different space regions. This is achieved in HRBF through stacking non-complete grids over a first grid at a coarse scale. The first grid will output a rough estimate of the surface, $a_1(P)$ (Fig. 2a) as:

$$a_1(P) = \sum_{k=1}^{M_1} \tilde{z}(P_{k1}) G(P; P_{k1} \mid \sigma_1) \Delta P_1^{\,2} \qquad (6)$$

The residual $\{r_1(P_j)\}$ is computed for each sampled data point, $(P_j, z(P_j))$, as:

$$r_1(P_j) = a_1(P_j) - z(P_j) \qquad (7)$$

This residual will be the input to a second grid which features half the scale of the first grid: $\sigma_2 = \sigma_1/2$. This second grid does not need to reconstruct the original surface but only the residual surface. Its reconstruction, $a_2$, will be at the scale $\sigma_2$, and it will provide a second residual $r_2$. This grid will not be full, but Gaussians will be inserted only when a poor approximation is given. This is

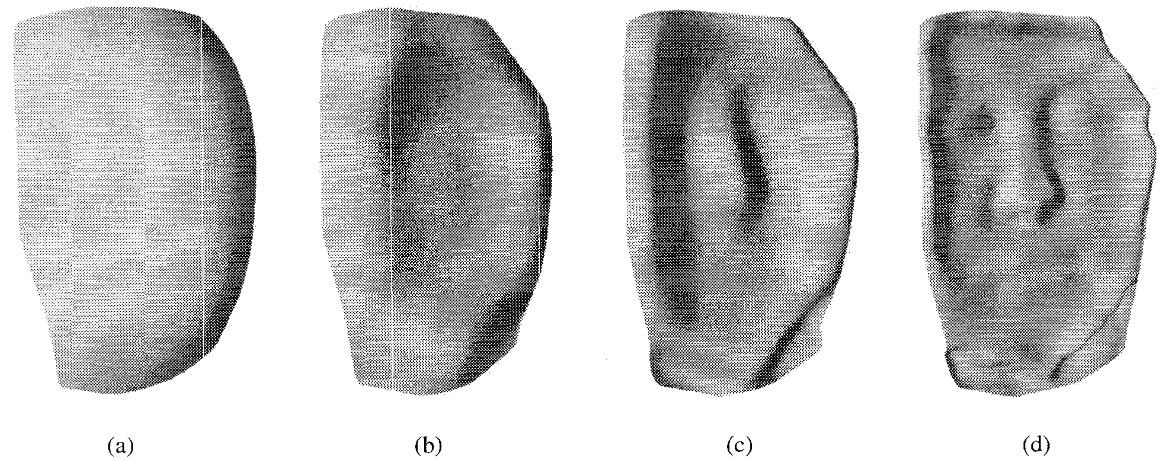


***Figure 2. The output of the four grids of the HRBF model.***

evaluated through the residual itself: a Gaussian is inserted in the grid crossing *kl* only if:

$$\frac{\sum_{P_r \in A(P_{kl})} \left| r_1(P_r) \right|}{N_k} < \varepsilon_n \qquad (8)$$

where $\varepsilon_n$ is the rms value of the noise and $N_k$ is the number of sampled points which belong to $A(P_{kl})$.

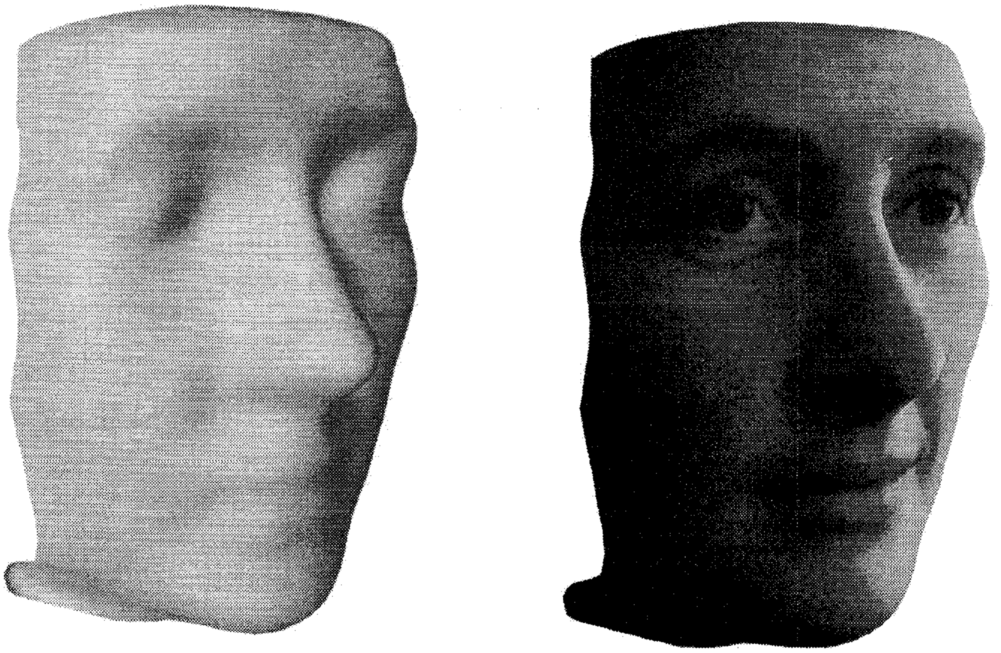

(a) (b)

**Figure 3. (a) The 3D model constructed adding the contribution of the four grids showed in Figure 2a-d. (b) The final result obtained mapping the texture on the model. It should be compared with Figure 1c.**

Grids are created one after the other, until condition (8) is not satisfied over the entire input domain (Figs. 2a-2d). Usually three-four grids are sufficient for face reconstruction. The final result is a uniform approximation of the surface in a $L^1$ metric. If a uniform convergence in a different metric were required, this can be achieved by simply changing the metric in (8).

The quantitative results are summarised in Table I. As can be seen from the second column the residual error decreases with the number of grids and it is essentially zero mean when four grids have been used as it is expected when only the measurement noise is left. Its standard deviation decreases towards 1 mm (third column). Although the digitisation noise is of one order of magnitude less, motion of the head, which is not constrained during the acquisition, makes the instrument less precise. A suitable chair would increase the accuracy of one order of magnitude. However, this would limit the acquisition freedom and flexibility, and it has not been considered here. The reconstruction implements a total of 6,986 Gaussians units which are much less than the 9,200 required by the full grid at the lowest scale (column 4). Moreover, the units are clustered in those regions where the highest details are found.

Finally, the obtained 3D surface can now be aligned with the bitmap obtained from the snapshot and the face texture can be mapped over it. The final result is a 3D very realistic reconstruction of the face (Figs 1c and 3b).

## 4 . References

[1] N.A. Borghese, G. Ferrigno, G. Baroni, R. Savarè, S. Ferrari and A. Pedotti, "AUTOSCAN: A flexible and portable scanner of 3D surfaces," *IEEE Computer Graphics & Applications*, vol. 18, no. 3, pp. 2-5, May/June 1998.

[2] N.A. Borghese, S. Ferrari, "Hierarchical RBF networks and local parameters estimate," *Neurocomputing*, vol. 19, pp. 259-283, 1998.

[3] M. Cannon & J.E. Slotine, "Space-Frequency localized basis function networks for nonlinear system estimation and control," *Neurocomputing*, vol. 9, pp. 293-342, 1995.

[4] P. Cignoni, C. Montani, E. Puppo and R. Scopigno, "Multiresolution Representation and Visualization of Volume Data," *IEEE Trans. Visualization and Computer Graphics*, vol. 3, no. 4, pp. 352-369, 1997.

[5] G., Ferrigno, and A. Pedotti. "Modularly expansible system for real-time processing of a TV display, useful in particular for the acquisition of coordinates of known shapes objects," U.S. patent 4,706,296, 1990.

[6] H. Hoppe, "Surface reconstruction from unorganized points," *PhD Thesis*, Dept. of Computer Science and Engineering, Univ. of Washington, June 1994.

[7] J. Moody and C. Darken, "Fast-learning in networks of locally-tuned processing units," *Neural Computation*, vol. 1, no. 2, 281-294, 1989.

[8] M. Petrov, A. Talapov, T. Robertson, A. Lebedev, A. Zhilayaev and L. Polonskiy, "Optical 3D Digitisers: Bringing Life to the Virtual World," *IEEE Computer Graphics*, vol. 18, no. 3, pp. 28-37, 1998.

[9] J. Platt, "A Resource-Allocating Network for Function Interpolation," *Neural Computation*, vol. 3, pp. 213-225, 1991.

[10] T. Poggio and F. Girosi, "Regularization Algorithms for learning that are equivalent to multilayer networks," *Science*, vol. 247, pp. 978-981, 1990.

[11] H. Urban, "Ultrasonic Imaging for industrial scene analysis," *Sensor Devices and System for Robotics*, NATO ASI series, Computer and System Sciences, Alicia Casals (ed.), vol. 52, pp. 187-194, Springler Verlag, 1989.

***Table I: Reconstruction of the 3D face reported in Figure 1***

| #grid | MSE | Mean Error | Error std | #gauss | $\nu_{cut\text{-}off}$ | $\sigma$ |
|---|---|---|---|---|---|---|
| | [mm$^2$] | [mm] | [mm] | | [Hz] | [mm] |
| 1 | 160.92 | 7.59 | 10.17 | 116/150 | 0.01 | 18.74 |
| 2 | 25.36 | 0.73 | 4.98 | 476/580 | 0.02 | 9.37 |
| 3 | 4.65 | 0.15 | 2.15 | 1812/2320 | 0.04 | 4.68 |
| 4 | 1.28 | 0.02 | 1.13 | 4582/9200 | 0.08 | 2.34 |